GENE CLONING

BACKGROUND OF THE INVENTION

1. FIELD OF THE INVENTION

This invention relates generally to methods and materials for use in gene cloning. More specifically, the present invention relates to gene probes/primers for use in discovery and characterization of bioactive compound coding genes and gene clusters.

2. DESCRIPTION OF RELATED ART

The basic challenges in drug discovery are to identify a lead compound with desirable activity, and to optimize the lead compound to meet criteria required to proceed with further drug development. One common approach to drug discovery involves presenting macromolecules implicated in causing a disease (disease targets) in bioassays in which potential drug candidates are tested for therapeutic activity. Such molecules could be receptors, enzymes or transcription factors.

Another approach involves presenting whole cells or organisms that are representative of the causative agent of the disease. Such agents include bacteria and tumor cell lines.

Traditionally, there are two sources of potential drug candidates: collections of natural products and synthetic chemicals. Identification of lead compounds has been achieved by random screening of such collections which encompass as broad a range of structural types as possible. The recent development of synthetic combinatorial chemical libraries will further increase the number and variety of compounds available for screening. However, the diversity in any synthetic chemical library is limited to human imagination and skills of synthesis.

Random screening of natural products from sources such as terrestrial bacteria, fungi, invertebrates and plants has resulted in the discovery of many important drugs (Franco et al. 1991, Critical Rev Biotechnol 11:193-276; Goodfellow et al. 1989, in "Microbial Products: New Approaches", Cambridge University Press, pp. 343-383; Berdy 1974, Adv Appl Microbiol 18:309-406; Suffness et al. 1988, in Biomedical Importance of Marine Organisms, D. G. Fautin, California Academy of Sciences, pages 151-157). More than 10,000 of these natural products are biologically active and at least 100 of these are currently in use spanning the entire therapeutic spectrum, including antibiotics, anti-cancer agents, and cardiovascular agents, and also as agrochemicals . The success of this approach of drug discovery depends heavily on how many compounds enter a screening program and how efficiently the screening can be conducted. Thus, indication-specific compound libraries have tremendous advantages to this end.

Typically, pharmaceutical companies screen compound collections containing hundreds of thousands of natural and synthetic compounds. However, the ratio of novel to previously-discovered compounds has diminished with time. In screens for anti-cancer agents, for example, most of the microbial species which are biologically active can yield compounds that are already characterized. This is due partly to the difficulties of consistently

and adequately finding, reproducing and supplying novel natural product samples. Since biological diversity is largely due to underlying molecular diversity, there is insufficient biological diversity in the organisms currently selected for random screening, which reduces the probability that novel

5   compounds will be isolated.

Novel bioactivity has consistently been found in various natural sources. See for example, Cragg et al., 1994. (in "Enthnobotany and the search for new drugs" Wiley, Chichester. p178-196). Few of these sources have been explored systematically and thoroughly for novel drug leads. For

10   example, it has been estimated that only 5000 plant species have been studied exhaustively for possible medical use. This is a minor fraction of the estimated total of 250,000-3,000,000 species, most of which grow in the tropics (Abelson 1990, Science 247:513). Moreover, out of the estimated millions of species of marine microorganisms, only a small number have been

15   characterized. Indeed, there is tremendous biodiversity that remains untapped as sources of lead compounds. Conventional methods of compound discovery from these sources is requisite on the successful laboratory culture of the microbial flora, a practice that is only approximately 1% efficient. Thus, the vast majority of environmental microorganisms cannot be grown in a

20   laboratory and therefore, any potential bioactive compounds that they produce cannot be assayed.

Terrestrial microorganisms, fungi, invertebrates and plants have historically been used as sources of natural products. However, apart from several well-studied groups of organisms, such as the actinomycetes, which

25   have been developed for drug screening and commercial production, reproducibility and production problems still exist. For example, the antitumor agent, TAXOL^{TM}, is a constituent of the bark of mature Pacific yew trees, and its supply as a clinical agent has caused concern about damage to the local

ecological system. Taxol contains 11 chiral centers with 2048 possible diastereoisomeric forms so that its *de novo* synthesis on a commercial scale seems unlikely (Phillipson, 1994, Trans Royal Soc Trop Med Hyg 88 Supp 1:17-19).

Marine invertebrates are a promising source of novel compounds but there exist major weaknesses in the technology for conducting drug screens and large-scale resupply. For instance, marine invertebrates can be difficult to recollect, and many have seasonal variability in natural product content.

Marine microorganisms are a promising source of novel compounds but there also exist major weaknesses in the technology for conducting drug screens and industrial fermentation with marine microorganisms. For instance, marine microorganisms are difficult to collect, establish and maintain in culture, and many have specialized nutrient requirements. A reliable source of unpolluted seawater is generally essential for fermentation. It is estimated that at least 99% of marine bacteria species do not survive on laboratory media. Furthermore, available commercial fermentation equipment is not optimal for use in saline conditions, or under high pressure.

Certain compounds appear in nature only when specific organisms interact with each other and the environment. Pathogens can alter plant gene expression and trigger synthesis of compounds, such as phytoalexins, that enable the plant to resist attack. For example, the wild tobacco plant Nicotiana sylvestris increases its synthesis of alkaloids when under attack from larvae of Manduca sexta. Likewise fungi can respond to phytoalexins by detoxification or preventing their accumulation. Such metabolites will be missed by traditional high-throughput screens, which do not evaluate a fungus together with its plant host. A dramatic example of the influence of the natural environment on an organism is seen with the poison dart frog. While a lethal dose of the sodium channel agonist alkaloid, batrachotoxin, can be harvested

by rubbing the tip of a blow dart across the glandular back of a field specimen, batrachotoxin could not be detected in second generation terrarium-reared frogs (Daly, 1995, Proc. Natl. Acad. Sci. 92:9-13). If only traditional drug screening technologies are applied, potentially valuable molecules such as these can never be discovered. Additionally, plant and vertebrate microbial symbionts can sometimes independently biosynthesize bioactive compounds originally discovered from the host plant. In fact, in many cases (e.g. taxanes), the symbiont population produces a much wider range of related compounds. It is believed that similar biosynthetic pathways exist in both host and symbiont as a result of horizontal gene transfer. Thus, symbiont microorganisms represent a virtually untapped source of novel natural products, but only if new methods are made available that can overcome the limitations of conventional methods in fermentation/culturing and discovery of compounds from environmental microorganisms.

Moreover, a lead compound discovered through random screening rarely becomes a drug, since its potency, selectivity, bioavailability or stability may not be adequate. Typically, a certain quantity of the lead compound is required so that it can be modified structurally to improve its initial activity. However, current methods for synthesis and development of lead compounds from natural sources, especially plants, are relatively inefficient. There are significant obstacles associated with various stages of drug development, such as recollection, growth of the drug-producing organism, dereplication, strain improvement, media improvement, and scale-up production. These problems delay clinical testing of new compounds and affect the economics of using these new sources of drug leads.

At present, the above-mentioned marine, botanical and animal sources of natural products are underused. Currently available methods for producing and screening lead compounds cannot be applied efficiently to these under-

explored sources. Unlike some terrestrial bacteria and fungi, these drug-producing organisms are not readily amenable to industrial fermentation technologies. Simultaneously, the pressure for finding novel sources for drugs

5 is intensified by new high-efficiency and high-throughput screening technologies. Therefore, there is a general need for methods of harnessing the genetic resources and chemical diversity of these as yet untapped sources of compounds for the purpose of drug discovery. Discovery through microbial symbionts offers one possibility if methods can be developed that overcome limitations inherent in conventional discovery from environmental

10 microorganisms.

Most recent drug discovery programs have shifted to mechanism-based discovery screens. Once a molecular target is identified (e.g., a hormone receptor involved in regulating the disease), assays are designed to identify and/or synthesize therapeutic agents that interact at a molecular level

15 with the target.

Gene expression libraries are used to identify, investigate and produce the target molecules. Expression cloning has become a conventional method for obtaining the target gene encoding a single protein without knowing the protein's physical properties.

20 Many proteins identified by screening gene expression libraries prepared from human and mammalian tissues are potential disease targets, e.g., receptors (Simonsen et al. 1994, Trends Pharmacol Sci 15:437-441; Nakayama et al. 1992, Curr Opin Biotechnol 3:497-505; Aruffo, 1991, Curr Opin Biotechnol, 2:735-741), and signal-transducing proteins. See Seed et

25 al., 1987, Proc Natl Acad Sci 84:3365-3369; Yamasaki et al., 1988, Science 241:825-828; and Lin et al., 1992, Cell 68:775-785, (type III TGF-ß receptor) for examples of proteins identified by functional expression cloning in mammalian cells.

Once a disease target is identified, the protein target or engineered host cells that express the protein target have been used in biological assays to screen for lead compounds (Luyten et al. 1993, Trends Biotechnol 11:247-54). Thus, within the scheme of drug discovery, the use of gene expression libraries has been largely limited to the identification and production of potential protein disease targets. Only in those instances where the drug is a protein or small peptide, e.g., antibodies, have expression libraries been prepared in order to generate and screen for molecules having the desirable biological activity (Huse et al. 1991, Ciba Foundation Symp 159:91-102).

However, there are other applications of gene expression libraries that are relevant to drug discovery. Gene libraries of microorganisms have been prepared for the purpose of identifying genes involved in biosynthetic pathways that produce medicinally-active metabolites and specialty chemicals. These pathways require multiple proteins (specifically, enzymes), entailing greater complexity than the single proteins used as drug targets. For example, genes encoding pathways of bacterial polyketide synthases (PKSs) were identified by screening gene libraries of the organism (Malpartida et al. 1984, Nature 309:462; Donadio et al. 1991, Science 252:675-679). PKSs catalyze multiple steps of the biosynthesis of polyketides, an important class of therapeutic compounds, and control the structural diversity of the polyketides produced. A host-vector system in Streptomyces has been developed that allows directed mutation and expression of cloned PKS genes (McDaniel et al. 1993, Science 262:1546-1550; Kao et al. 1994, Science 265:509-512). This specific host-vector system has been used to develop more efficient ways of producing polyketides, and to rationally develop novel polyketides (Khosla et al., WO 95/08548).

Another example is the production of the textile dye, indigo, by fermentation in an E. coli host. Two operons containing the genes that encode the multienzyme biosynthetic pathway have been genetically manipulated to improve production of indigo by the foreign E. coli host.(Ensley et al. 1983, Science 222:167-169; Murdock et al. 1993, Bio/Technology 11:381-386). Overall, conventional studies of heterologous expression of genes encoding a metabolic pathway involve cloning, sequence analysis, designed mutations, and rearrangement of specific genes that encode proteins known to be involved in previously characterized metabolic pathways.

In view of numerous advances in the understanding of disease mechanisms and identification of drug targets, there is an increasing need for innovative strategies and methods for rapidly identifying lead compounds and channeling them toward clinical testing.

The speed and availability of automated nucleic acid synthesis has led to rapid technological advances in biological research. For example, the availability of synthetic primers for sequencing has permitted researchers to decrease their time and labor involved in sequencing a particular nucleic acid by approximately sixty percent. Another technology which is facilitated by synthetic oligonucleotides is the polymerase chain reaction (PCR). This technique, which involves the exponential amplification of sequences between two synthetic primers, offers unprecedented detection levels and permits genetic manipulation of the amplified sequence. Further, the availability of synthetic primers allows a variety of genetic manipulations to be performed with relatively simple procedures, including site-specific mutagenesis and the custom design of genetic vectors.

Sequences to be cloned are also routinely modified with synthetic oligonucleotides. The modifications of either vector or insert sequence can range from the addition of a simple sequence encoding a restriction enzyme

site to more complicated schemes involving modifying the translation product of the cloned sequence with a specific peptide or a variety of peptide sequences. Thus, these technological advances associated with synthetic oligonucleotides has afforded researchers many opportunities to study diverse biological phenomenon in greater detail and with greater speed and accuracy.

Oligonucleotide synthesis proceeds via linear coupling of individual monomers in a stepwise reaction. The reactions are generally performed on a solid phase support by first coupling the 3' end of the first monomer to the support. The second monomer is added to the 5' end of the first monomer in a condensation reaction to yield a dinucleotide coupled to the solid support. At the end of each coupling reaction, the by-products and unreacted, free monomers are washed away so that the starting material for the next round of synthesis is the pure oligonucleotide attached to the support. In this reaction scheme, the stepwise addition of individual monomers to a single, growing end of a oligonucleotide ensures accurate synthesis of the desired sequence. Moreover, unwanted side reactions are eliminated, such as the condensation of two oligonucleotides, resulting in high product yields.

In some instances, it is desired that synthetic oligonucleotides have random nucleotide sequences. This result can be accomplished by adding equal proportions of all four nucleotides in the monomer coupling reactions, leading to the random incorporation of all nucleotides and yields a population of oligonucleotides with random sequences. Since all possible combinations of nucleotide sequences are represented within the population, all possible codon triplets will also be represented. If the objective is ultimately to generate random peptide products, this approach has a severe limitation because the random codons synthesized will bias the amino acids incorporated during translation of the DNA by the cell into polypeptides.

The bias is due to the redundancy of the genetic code. There are four nucleotide monomers which leads to sixty-four possible triplet codons. With only twenty amino acids to specify, many of the amino acids are encoded by multiple codons. Therefore, a population of oligonucleotides synthesized by sequential addition of monomers from a random population will not encode peptides whose amino acid sequence represents all possible combinations of the twenty different amino acids in equal proportions. That is, the frequency of amino acids incorporated into polypeptides will be biased toward those amino acids which are specified by multiple codons.

To alleviate amino acid bias due to the redundancy of the genetic code, the oligonucleotides can be synthesized from nucleotide triplets. Here, a triplet coding for each of the twenty amino acids is synthesized from individual monomers. Once synthesized, the triplets are used in the coupling reactions instead of individual monomers. By mixing equal proportions of the triplets, synthesis of oligonucleotides with random codons can be accomplished. However, the cost of synthesis from such triplets far exceeds that of synthesis from individual monomers because triplets are not commercially available.

It would therefore be useful to develop a method for synthesizing oligonucleotides which are designed for hybridizing to genes coding for bioactive compound coding genes, antibiotics, and secondary metabolites.. The present invention satisfies these needs and provides additional advantages as well.

SUMMARY OF THE INVENTION

According to the present invention, there is provided a method of targeted cloning and enrichment of genes and gene clusters. This is accomplished by directly cloning the target gene from the source DNA using one of several novel methods presented, for example by creating template derived primers containing target oligonucleotides, adding these template derived primers to a sample of DNA and performing PCR to replicate those genes targeted by the template derived primers. The methods provide the degenerate cloning of the entire family of related target genes from a mixed DNA sample. This collection of related genes is then used to affinity purify and clone larger target gene containing fragments from the sample, representing associated biosynthetic pathway genes. The result is a target gene/pathway enriched genomic library. Also provided are the genes provided by this method and the probes used in connection with this method. These are also useful for hybridization screening of clonal libraries as well as culture collections.

DESCRIPTION OF THE DRAWINGS

Other advantages of the present invention will be readily appreciated as the same becomes better understood by reference to the following detailed description when considered in connection with the accompanying drawings wherein:

Figure 1 is a photograph showing a gel of the results of the degenerate nested pair PCR reaction for cloning the DHFR2 gene probe from marine sediment DNA; Lanes 10-15 are the products from the first PCR using

primers DHFR2-1 and DHFR2-4; Lanes 3-8 are the products from the second PCR using products from the first PCR as template and primers DHFR2-2 and DHFR2-3; Lane 9 contains size markers; Reaction conditions were as specified in the text; The expected product size is about 120 bp as seen in

5   lanes 8 and 4;

Figures 2 A, B, and C illustrate the strategy for generating template specific primers and their use in specific cloning of unknown flanking sequences against a single known primer, details are discussed in the text;

10

Figure 3 shows the PCR amplification of the part of Amp resistant gene from pBR325 as template using BcgI derived primers and sequence specific pPstCW primer, the reaction mixture contains pBR325 digested with BamHI (50ng) as template, 32-mer BcgI primers from pBR325 (gel purified 12pmol)

15   and/or 32-mer sequence-specific primer pPstCW (40pmol); BcgI primers were denatured five minutes 100 C and immediately cool down on ice; For PCR used program AF08 (T1=96 C, t=30 seconds; T2 56 C, t=1 minutes; T3=72 C, t=10 seconds, reactions are carried out in 34 cycles); Lanes: 1. Mixture of BcgI oligonucleotide with template; 2. Mixture of BcgI

20   oligonucleotides plus sequence specific pPstCW primer with template; 3. pPstCW primer with template but without BcgI primers; 4. pPstCW primer plus BcgI primers mixture without template; and 5. pBR325 (1ug) digested with BcgI restriction endonuclease;

25   Figure 4 shows the purified single-stranded DNA form from phages run on 0.8% agarose gel (ethidium bromide stained); Lanes: 1. SS DNA M13mp18 (1ug); 2. SS DNA from M13mp18 BcgI library from *E. coli* HB101 (1ug); 3. SS DNA from M13 mp18 BcgI library from *S. clavuligerus* (1ug); and

4. DNA marker lambda DNA digested HindIII (1ug) (Promega, WI);

Figure 5 shows a 0.8% agarose gel analysis of biotinylated polymerase elongation products (BEPEP); Panels: **A** ethidium bromide stained gel. **B** steptavidin-phosphatase assay from southern blotting. PCR and polymerase elongation reaction (PER) carried out in 20 ul reaction format contained 30pmol biotinylated primers and 100ng for PCR or 1ug for PER DNA template either CsCl purified DNA from *E. coli* HB101 or DNA from *S. clavuligerus*; For improving elongation, reactions carried out with 5u TaKaRa LA DNA polymerase according to the manufacturer's instructions (TaKaRa, Japan); Primers were labeled with photobiotin (Vector, CA) according manual; Lanes: 1. PCR from HB101 DNA, primers AA and AB (sequence-specific primers for PEPase gene of *E. coli*), 2. PCR from *S. clavuligerus* DNA with primers ACVS 010 and 011 (ACVS 04-011 are sequence-specific primers for ACVS gene of *S. clavuligerus*); 3. Negative control PCR primers AA and AB without template; 4. biotinylated DNA marker lambda digested HindIII (0.5 ug); 5. Negative control PCR primers ACVS 010 and 011 without template; 6. BPEP with AA primer and DNA HB101; 7. BPEP with AB primer and DNA HB101; 8. BPEP with ACVS 04 primer and DNA *S. clavuligerus*; 9. BPEP with ACVS05 and DNA *S. clavuligerus*; 10. BPEP with ACVS08 and DNA *S. clavuligerus*; 11 BPEP with ACVS 09 and DNA *S. clavuligerus*; 12. BPEP with ACVS 010 and DNA *S. clavuligerus*; and 13. BPEP with ACVS 011 and DNA *S. clavuligerus*;

Figure 6 shows a blot of a streptavidin-phosphatase assay of binding of biotinylated polymerase elongated products (BPEP) to Avidin D beads (Vector Labs, CA); Lanes : 1. BPEP with primers AA and AB from DNA *E. coli* HB101 of. 2. BPEP with primers ACVS 04-011 from *S. clavuligerus* DNA. Panels: **A** BPEP before adsorption to beads, B BPEP fraction unbound to

- 14 -

beads, C BPEP fraction incubated at 65 C in TBST buffer; 50 ul (25ng/ul) Avidin beads equilibrated two times with three volumes of TBST buffer and mixed with 100 ul BPEP and incubated 2 hours at 37°C; Then beads washed three times with three volumes of TBST buffer and 2ul analyzed on
5   streptavidin-phosphatase dot blotting assay;

Figure 7 shows a photograph of an agarose gel analysis of bound and unbound fractions of SS DNA form M13mp18 Bcgl libraries; Lanes: 1.,7. DNA marker lambda HindIII (1ug); 2. M13mp18 original Bcgl library from *E. coli*
10 HB101 (10ug); 3. Unbound fraction of HB101 Bcgl library at 37°C; 4. Bound fraction of Bcgl HB101 library at 37°C; 5. Unbound fraction of Bcgl HB101 library at 65°C; 6. Bound fraction of Bcgl HB101 library at 65°C; 8. M13mp18 original Bcgl library from *S. clavuligerus* (10ug); 9. Unbound fraction of Bcgl *S. clavuligerus* library at 37°C; 10. Bound fraction of Bcgl *S. clavuligerus*
15 library at 37°C; 11. Unbound fraction of Bcgl *S. clavuligerus* library at 65°C; 12. Bound fraction of Bcgl *S. clavuligerus* library at 65°C,10 ul (1ug/ul) of M13 mp18 SS DNA library mixed with 50 ul of Avidin D-BPEP (biotinylated polymerase elongated product) in 100 ul of TBST buffer and incubate overnight at 55°C; The temperature was then decreased to 37°C for 10
20 minutes and 100ul of unbound fraction was collected, The beads were washed three times with three volumes of TBST buffer for five minutes, bound fraction was eluted with 100ul of water by boiling at 100°C for five minutes; All fractions were ethanol precipitated and dissolved in 30 ul water; 10 ul was analyzed by agarose gel electrophoresis and 3ul was
25 electrotransformed into Nova Blue *E. coli* cells (Novagene, WI);

Figure 8 is a photograph of an agarose gel used to analyze the PCR amplification of *S. clavuligerus* DNA with sequence-specific ACVS and

captured BcgI primers; PCR reactions carried out in 20 ul format with 100ng of *S. clavuligerus* genomic DNA as template and 30 pmol primers; Lanes: 1. ACVS 04 plus 011; 2. ACVS 09 plus 011; 3. ACVS 08 plus 010; 4. ACVS 010 plus 011; 5. DNA marker lambda HindIII; 6. ACVS 04 plus 04w3bcg; 7. ACVS 04 plus 04w6bcg; 8. ACVS 04 plus 04w9bcg; 9. ACVS 04 plus 04w10bcg; 10. ACVS04 plus 04w13bcg; 11. ACVS09 plus 04w3bcg; 12. ACVS09 plus 04w6bcg; 13. ACVS09 plus 04 w9bcg; 14. ACVS09 plus 04w10bcg.; and 15. ACVS09 plus 04w13bcg;

Figure 9 is a photograph of an agarose gel showing the PCR amplification products of using octamer primers calculated using the k-tuple strategy as described in the text; Template used is HB101 genomic DNA and otherwise standard conditions; Lanes 1, 1', 18' contain size markers; 2-9, oct03 as a solitary primer with varying buffer compositions (lanes 10-18 are empty); 2'-9' standard primers for the phosphoenol pyruvate gene as a control; 10'-18' oct01 as a solitary primer with the same varying buffer as in lanes 2-9 Products are of expected size for a random PCR (0.2-3 kb);

Figure 10 is a photograph of an agarose gel comparing the PCR amplification products from Figure 9; Lanes: 1, markers; 2, oct01; 3 oct03; Reactions were conducted under optimized conditions as judged from analysis of reactions shown in Figure 9;

Figure 11 is a photograph of an agarose gel showing the PCR products using k-tuple generated primers pair-wise with primers specific for the *acvs* gene; *S. clavuligerus* genomic DNA was used as template, under otherwise standard cycling conditions and temperature gradient (each primer pair PCR was conducted at 27°C, 34°C, and 42°C, left to right across the gel); Lanes:

1, size markers; 2-4, ACVS05 and oct01; 5-7, ACVS05 and oct02; 8-10 ACVS07 and oct01; 11-13, ACVS07 and oct02; Controls confirmed that amplification was due to pair-wise priming of specific and octamer primers, and not solitary priming by either primer alone;

Figure 12 is a photograph of a hybridization blot analysis of a streptavidin-phosphatase assay of different fractions of biotinylated PCR probes during purification on Avidin DLA beads; Panels: **A)** original mixture of PCR probes 2ul (50ng/ul); **B)** unbound fraction of non-biotinylated PCR probes 2ul (50ng/ul); **C)** biotin eluted fractions of biotinylated PCR probes 2ul (10ng/ul); Lanes: 1. Bio IPNS 05+06 PCR product; 2. Bio StsC03+04 PCR product; 200ul of Avidin DLA beads were used for purification (capacity 25ng/ul);

Figure 13 is a photograph showing a 1% agarose gel electrophoresis analysis of Avidin DLA purified biotinylated PCR probes; Lanes: 1. Bio IPNS 05+06 5ul (10ng/ul); 2. Bio StsC 03+04 5ul (10ng/ul); Panels: **A)** ethidium bromide stained gel; **B)** streptavidin-phosphatase assay from the southern blotting of the gel;

Figure 14 is a photograph of the results of screening pFD666 and pSCOS1 S. griseus genomic libraries, enriched for aminoglycoside genes, with alk-direct labeled StsC03+04 probes; Panels: **A)** original library (total of 500 colonies on the plate); **B)** StsC and recA captured library after eletrotransformation (total 250 colonies on plate); **C)** library derived from StsC and recA captured chromosomal DNA fragments cloned into pSCOS1 cosmid vector (total 2000 colonies on plate); Results demonstrate over a 100-fold enrichment for the specific gene, as compared to the expected number of

positive clones in the unenriched library;

Figure 15 is a photograph of several dot-blots of positive clones from libraries enriched for the acvs gene (left panel) and strB1 (right panel), corresponding to the beta lactam and aminoglycoside biosynthetic clusters, respectively; Genomic libraries were constructed from S. clavuligerus (acvs) and S. griseus (strB1); DNA from positive clones frequently hybridized with several additional gene probes associated with their respective clusters (Table II), demonstrating the cloning of intact clusters and entire pathways;

Figure 16 is a photograph of an agarose gel used in the PCR analysis of several clones enriched for the aminoglycoside cluster (Figure 14); Lane 1 to 30: 1st PCR, using E. coli cells with different plasmids as template; Lane 1' to 30': 2nd PCR, using 1st PCR products as template; 1: MW marker (100bp ladder); 2 to 8: 1st PCR using StrD primers; 9 to 15: 1st PCR using StsA primers; 16: MW marker (100bp ladder); 17 to 23: 1st PCR using StrB1 primers; 24 to 30: 1st PCR using StsC primers; 1': MW marker (100bp ladder); 2' to 8': 2nd PCR using StrD primers; 9' to 15': 2nd PCR using StsA primers; 16': MW marker (100bp ladder); 17' to 23': 2nd PCR using StrB1 primers; 24' to 30': 2nd PCR using StsC primers, each set of seven lanes with same primer follows the same pattern of order: (1) no template; (2) PDF666 as template; (3) B1-1 as template; (4) B3-1as template; (5) B20-2 as template; (6) B20-4 as template; (7) B16str5 as template; These results confirm the retention and stable cloning of the cluster in many clones and corroborates the hybridization results indicating the presence of these genes; Additionally, the utility of many of the oligos listed in Table I and used in double nested PCR as described herein is also demonstrated;

Figure 17 is a photograph of antibiotic selection plates demonstrating the heterologous expression of *S. griseus* aminoglycoside resistance in *E. coli*;   Left Panel: gradient plates from 0 (bottom) to 25 (top) ug/ml streptomycin;   The left side of each plate contains a spread from a single positive colony that hybridized to the strB1 gene probe; the right side of each plate contains the *E. coli* host transformed with cosmid containing no insert; Right Panel: plates contain the same clones as in the left panel;   Plates contain 0, 5, 15, 25 ug/ml streptomycin, clockwise from the upper left plate; and

Figure 18 shows yet another example of hybridization probing of genomic libraries using several of the gene probes;   SFT4 is a library constructed from a trimethoprim resistant seawater isolate, carries the DHFR2 gene, and demonstrates antibacterial activity against *S. aureus* in standard antibiotic challenge assays.

DETAILED DESCRIPTION OF THE INVENTION

Generally, the present invention provides a method and probes for use in targeted cloning and enrichment of genes and gene clusters from an otherwise mixed and very diverse population of DNA. The methods provide the degenerate cloning of the entire family of related target genes from a mixed DNA sample. This collection of related genes is then used to affinity purify and clone larger target gene containing fragments from the sample, representing associated biosynthetic pathway genes. The result is a target gene/pathway enriched genomic library.   Also provided are the genes provided by this method and the probes used in connection with this method.

These are also useful for hybridization screening of clonal libraries as well as culture collections.

Genomics and bioinformatics can be used to identify specific genes and DNA sequences that correlate with the biosynthesis of specific structural classes of compounds, including many secondary metabolites. This is often conducted through a comparison of either the nucleotide gene sequences of known related genes or the protein sequences of the gene products through multiple sequence alignments. Constant or conserved regions within related sequences are thought to be important for protein function and will also be conserved in undiscovered genes of the related class. Cloning the entire population of target genes coding for a specific function allows for the associated, clustered biosynthetic pathways to also be cloned in a very specific and targeted manner (see below). Additionally, using degenerate PCR cloning permits the cloning of both closely as well as distantly related genes within a specific target class, subsequently permitting the cloning and capture of the entire genetic and chemical diversity for the target compound class of interest.

Degenerate-nested temperature gradient PCR is used for the successful cloning of the majority or even entire population of related genes from a mixture of many genomes and otherwise unrelated DNA, such as the total DNA isolated from a sample of soil or other environmental source. Nested sets of degenerate PCR primers have been designed for a variety of target genes (see TABLE I).

Several oligonucleotide PCR primers and hybridization probes were designed and then synthesized to target DNA sequences from a variety of sources that potentially contain bioactive compound coding or resistance genes. The design of each oligo was conducted based on the alignment of sequences of the gene and/or protein family of interest that are available

publicly (i.e. through GenBank). Several sequences were used, if available. In some cases, only a single unique sequence was available and used in calculating the oligo sequences. Most oligos were designed as degenerate nested pairs in order to maximize their capacity for the cloning and discovery
5   of both closely, as well as distantly related novel sequences that, likewise, code for novel proteins and enzymatic products, such as secondary metabolites useful as lead drug compounds for screening.

The general method used for cloning target genes using degenerate nested temperature gradient PCR uses the following steps. First, a
10 temperature gradient for the 1st PCR is established having a range of temperatures from 41-60°C. This is accomplished using types of buffers having a pH of 8.3-9.2, $MgCl_2$ (1.5-3.5 mM), KCl (25 & 75mM) (Stratagene PCR Optimization Kit). A volume range between 10-30ul per reaction is placed in a 0.2ml tube for cycles between 30-35. This is ten times diluted and
15 the 1st PCR products are used as templates for a 2nd PCR reaction. The 2nd PCR occurs at 52°C and the other conditions are same with 1st PCR. A gel the then run and expected size of product is cut. DNA is extracted from the gel by using gel extraction kit (Qiagen). The PCR product is cloned into a pT7Blue-3 vector (Novagen), based on manufacturer's protocol. Clones
20 containing the target PCR product are screened by PCR and/or dot-blot hybridization. The plasmid is then purified. Automated sequencing is done using a Thermo Sequenase Cy5.5 terminator cycle sequencing kit (Amersham) and 50-250 fmol template with M13 forward or M13 reverse primers (2 pmol each).
25 The new sequence is aligned and compared with consensus target genes to confirm the degree of uniqueness by performing a BLAST search and sequence analysis.

In all cases, using these degenerate primers in the following way improves amplification significantly and reduces the number of unrelated misprimed products. This is a problem when it is otherwise desirous to sequence directly PCR products in the discovery of new genes. Most
5   misprimed products can be eliminated by conducting a degenerate, limited-degenerate, nested PCR (DLDN-PCR). The first PCR can be highly degenerate, which aids in the potential discovery of distantly related genes. However, this also results in more unrelated amplification products. The result is clearly seen on an agarose gel of the PCR reaction, where it is seen
10   that the expected product band is rather diffuse, and there is the existence of products of unexpected sizes (Figure 1, lanes 14, 15). Analysis of this band on a 10% acrylamide gel reveals the presence of several bands, each varying slightly in size, and presumably sequence. However, conducting a second PCR using the gel purified product band or first PCR reaction mixture as
15   template and a nested set of less degenerate primers results in amplification of only specific template targets (Figure 1, lanes 4, 8). In this case, a 150 bp product is clearly resolved and amenable for purification and sequencing. This is because the chances for an unrelated misprimed products also containing another mispriming site (in addition to the first which misprimed in
20   the first PCR) is very remote. This was confirmed by cloning and sequencing the PCR products from a single, yet diffuse DHFR2 band of about 150 bp. Although the cloned sequences all contained the priming sites, a much smaller percentage contained any related DHFR2 sequence bounded by the primer sites. Thus, a second PCR results in amplification of only the truly
25   related molecules from the population of products. Cloning and sequencing the products from this second PCR demonstrates this "filtering" effect. The result is a reliable strategy for generating degenerate PCR products amenable for direct sequencing. However, if the number of specific products

is high, as is sometimes the case for relatively common amplicons in environmental samples, then cloning the PCR products results in a large number of clones with specific product for sequencing.

Using the above described primers (Table I) and DLDN-PCR, there was discovered and sequenced several unique genes from marine and terrestrial microbial genomes related to a wide variety of biosynthetic pathways and structural classes of compounds, including antimetabolites, beta-lactams, polyketides, other antibiotics, taxanes, and others.

An example of this method involves a SC16RA01 probe which was generated using the "universal" 16S RNA PCR primers to amplify a 600 bp DNA product from S. clavuligerus. This probe is useful for colony hybridization probing for Streptomycetes and other related high GC content genomes. Additionally, this probe has been used in the PCR amplification of similar genomic DNA from a heterogeneous population.

The resulting gene probes can be used for the discovery of either single genes or entire clusters of adjacent genes involved in the total synthesis of compounds of interest, for example secondary metabolite biosynthetic pathways, the products of which comprise very useful libraries for antibiotic and other therapeutic compound screening. This is especially promising since the relatively recently emerging picture of the clustering of secondary metabolite gene pathways on the bacterial and fungal chromosome.

The following adaptation of the present invention describes a method for the generation and use of highly specific PCR primers derived from the template itself. However, their sequence need not be known *a priori*. This adaptation also exploits some unique and novel properties of restriction endonucleases, using BcgI as an example (Figure 2).

Bcgl is a novel Type II restriction endonuclease originally isolated from *Bacillus coagulans*, and is now commercially available. The recognition sequence for Bcgl is shown in the following and consists of a specific 6 base pair site of DNA sequence. However, the enzyme cleaves outside of this recognition site and generates a 32 bp restriction fragment:

$$5'-/(N)_{10}CGA(N)_6TGC(N)_{12}/-3'$$
$$3'-/(N)_{12}GCT(N)_6ACG(N)_{10}/-5'$$

Each restriction fragment is statistically unique in sequence and can be used as a specific oligonucleotide primer. The frequency of occurrence of the recognition site is the same as that for a random six base sequence, or about once every 4,000 nucleotides (i.e. $(1/4^6)$. However, the uniqueness of the fragment is extraordinary because it contains 34 nucleotides and corresponds to a randomized occurrence of once in 2.9×10^{20} bases. Random sequence analysis has confirmed the uniqueness of these restriction fragments and that they are not merely a frequently occurring repeat. This provides the basis for these fragments serving as very specific PCR primers and hybridization probes, each fragment highly specific for its own recognition sequence. By digesting an entire genome, entire or partial chromosome, or mixture of many genomes, very specific primers can be produced with priming sites spaced approximately 4,000 bp apart along the template DNA, ideal for PCR amplification and cloning.

The library of these unique oligonucleotides that are produced from strain specific genomic DNA or a mixed population of environmental DNA can be used as a set of primers for PCR and in combination with gene-specific primers, can be used for amplification and cloning of neighboring regions of DNA surrounding specific genes. Therefore, this technique can also be used for cloning large segments of DNA adjacent to a specific target site, including

complete bacterial operons, or biosynthetic pathway gene clusters from any organism. More broadly stated, this adaptation of the present invention can selectively and very efficiently (i.e. with high selectivity) amplify and clone from a mixture of DNA the regions flanking any specific target without any prior knowledge of the sequence to be cloned.

The simplest application of this adaptation of the present invention is to use the entire set of BcgI template derived oligonucleotide primers in a PCR that also contains a target specific oligonucleotide. A model system has been developed with pBR325. Then, the method of the present invention was used to amplify a 300 bp fragment of the ampicillin resistance gene using a specific primer and a random mixture of template derived primers from a BcgI digest of the pBR325 plasmid, which contains three BcgI cleavage sites (Figure 3). It was also determined that this method was effective with both linear and circularized template, using otherwise conventional PCR conditions.

An example of a more extensive and specific application of this adaptation of the present invention involves the identification and isolation from the entire BcgI restriction digest of the single oligonucleotide containing the priming site most proximal to a specific oligonucleotide on the template to be amplified and cloned. The following steps describe the method of the present invention:

1.   DNA isolation and purification from bacterial strains or total environmental DNA from sources such as oil, water, etc. using known procedures such as guanidine thiocyanate, CTAB, cesium chloride gradient or their combination and/or modification;

2.   Digestion of isolated DNA with BcgI endonuclease (NEB, protocol) and preparative purification of BcgI 34-mer oligonucleotides using 15% PAGE or 2% agarose gel in combination with the QIAEX II purification

system (Qiagen, CA) or any similar purification system.

3.   Construction of a 32-mer Bcgl oligonucleotide DNA library in M13 phage or any other phagmid vector, that does not contain Bcgl sites. Vector is first digested with SmaI, EcoRV or any other blunt-end producing unique restriction endonuclease followed by phosphatase (CIP) treatment. The purified 32-mer Bcgl oligonucleotides are treated with Klenow fragment of DNA polymerase I or T4-DNA polymerase in conjunction with polynucleotide kinase in the absence of any dNTP, but in the presence of ATP in order to convert 3'-protruding ends generated by Bcgl restriction endonuclease to blunt ends appropriate for cloning. The Bcgl restriction fragments are now 32 bp. Equimolar concentrations of the vector and blunt ended 32-mer oligonucleotides are ligated using T4 DNA ligase, followed by transformation into any conventional specific strain of *E. coli* (JM101, TG1 or ER2267) by chemical transformation or electroporation using conventional protocols;

4.   The library of phages is washed out from the agar plates following transformation and single stranded DNA is purified by standard methods (Figure 4).

5.   Specific primer (specific probe for the gene of interest) is labeled with biotin either at the 5'-end or randomly using a biotin labeling system (Vector Labs), or any other labeling system (e.g. fluorescein) .

6.   A single stranded, labeled copy of the sequence to be cloned is produced as follows. Annealing and elongation of the labeled, gene specific primer with genomic DNA template produces a single-stranded, biotinylated copy of the DNA of interest, including sequence downstream and flanking the known region, that which contains the annealing site of the gene specific probe (Figure 5). The biotinylated copy of DNA is isolated by absorption onto an avidin or streptavidin containing matrix, such as Avidex (Vector Labs, CA) or any other affinity matrix (Figure 6);

7.   Single stranded oligonucleotide DNA from the phage library (step 4) is then hybridized with single stranded biotinylated DNA under appropriate conditions. All non-specific phase DNA is then washed out and only phase DNA containing complementary sequences will hybridize to the biotinylated DNA. A subsequent boiling procedure releases the single-stranded phase DNA that can then be amplified via retransformation into *E. coli* and amplified *in vivo* (Figure 7).

8.   The phage library can be used for the generation of second primers either with PCR of the polylinker region or by BcgI digestion; Repeating steps 5-7 results in a nested set of PCR primers that can be used to amplify an entire biosynthetic pathway;

9.   The phage library is used for generation of second primers for PCR. Of many ways this can be accomplished, two examples were described. First, the 32 bp region of insert was sequenced directly and this sequence was used for oligonucleotide synthesis. As an example, this yielded several primers, including

GGGTCCGGCAGACCGTTCGCGGGCCGGAC,
GAGCGGACCGCACCGCGATCGGAACAACCT,
TCTCCGGGGCAGCGCGGTCGCGGAACGT.

A BLAST search confirmed their relatedness with the genus *Streptomyces*, as expected. Second, the polylinker region containing the 32 bp insert (desired PCR cloning primer) was amplified by PCR using the M13 universal and M13 reverse primers, generating a 184 bp PCR product. The 32 bp Bcg I fragment is flanked with unique EcoRI and BamHI restriction endonucleases sites and restriction with these enzymes was used to generate a 52 bp fragment, which was subsequently converted into a set of nested single-stranded

oligonucleotides by treatment with ExoIII nuclease under standard conditions. Each oligonucleotide in this nested set has the same 5'-end but a different level of deletion at 3'-end. Therefore, it can be used for PCR cloning as a second primer against a specific target primer. This approach can be used to clone several full length genes, operons, and entire biosynthetic pathways, as demonstrated in the next step.

10. This method was used to clone a region flanking a specific priming site within the *acvs* gene of *S. clavuligerus*. Combination of the first primer (gene specific) and secondary primers in a PCR results in the generation of sequences flanking that of the gene specific primer annealing site (FIGURE 8). Additional sequences can be subsequently cloned and combined into one operon for expression of proteins that produce secondary metabolites of interest (for example antibiotics).

Yet another adaptation of the present invention describes a method for the generation and use of highly specific PCR primers with frequently occurring priming sites across a wide range of genomes. These primers are novel and very useful for cloning sequences flanking a target sequence with no prior knowledge of the sequence to be cloned. This set of primers, collectively, is useful as a universal primer library, with specificity based on the criteria used in its generation.

This adaptation of the present invention relies on the analysis and interpretation of DNA sequences from a variety of genera, searching for relatively long and frequently repeated sequences across a wide range of genomes.

As an example, a genomic analysis of bacterial DNA protein coding regions was conducted and subsequently a set of 21 universal priming octamer oligonucleotides was constructed based on a very high frequency of repeating 8 base sequences. In addition, PCR conditions were optimized

using various thermopolymerases, including the Stoffel polymerase fragment, and have demonstrated the ability of these "universal primers" to prime against specific target primers. A 10 base universal oligonucleotide primer library was also constructed, and the sequence analysis data reveals that virtually any length oligonucleotide set can be constructed. However, the frequency of occurrence decreases with increasing oligonucleotide length. However, long amplification PCR techniques make even these highly specific but less frequently binding oligonucleotide quite useful.

As an example, octamer and decamer oligonucleotide libraries were generated by performing a k-tuple search and analysis using a proprietary gene database. This database consisted of 15 genera representing 34 bacterial and 4 fungal species, and 38 protein coding genes. The species included in this database were represented in a weighted fashion based on the known/perceived frequency and importance of secondary metabolite production. Of the nearly 65,000 octamers calculated, only a subset of approximately 200, or approximately 0.3%, were frequently present within every or most of the genes included in the database, and thus useful for universal PCR cloning. For example, the octamer OS-OCT-003 with sequence CTCGCCGA occurs 30 times and at least once in nearly every species. This corresponds to a determined average frequency of once every 1,625 nucleotides, while the random frequency for an eight base sequence is only once every 65,000 nucleotides. Similar calculations based on $k = 10$ resulted in a smaller number of equally frequent 10 base sequences, also useful for PCR primers. An example of 25 octamers and 12 decamers generated and used successfully for cloning are shown in Tables II and III

TABLE II

High Frequency Bacterial CDS Octamers

Name	Sequence, 5'—3'	Frequency
OS-OCT-001	GTCGGCGA	30
OS-OCT-002	CCAGATCG	21
OS-OCT-003	CTCGCCGA	23
OS-OCT-004	CGACATCG	18
OS-OCT-005	GCCGATCA	17
OS-OCT-006	GCCACCGA	15
OS-OCT-007	GATGCCGA	17
OS-OCT-008	CGGCGAAG	19
OS-OCT-009	CGGCGAAC	19
OS-OCT-010	GGCGATCA	15
OS-OCT-011	GCCGAGGA	17
OS-OCT-012	CGCCGACA	17
OS-OCT-013	ATCGCCGA	13
OS-OCT-014	GGCGAACC	13
OS-OCT-015	GCCGACCA	14
OS-OCT-016	GCCAAGGA	15
OS-OCT-017	CGGCAACG	16
OS-OCT-018	GGCTGGAC	13
OS-OCT-019	GCAGCACC	14

OS-OCT-020	CCAGCCAG	16
OS-OCT-21	CGCCGCCG	39
OS-OCT-22	CGGCGACC	34
OS-OCT-23	CCGCCGCC	33
OS-OCT-24	CGCGGCCG	31
OS-OCT-25	GTCGGCGA	30

TABLE III

High Frequency Bacterial CDS Decamers

Name	Sequence, 5'— 3'	Frequency
OS-DEC-001	CAGCTCGGCG	8
OS-DEC-002	GCCGGTGAGC	7
OS-DEC-003	CCGGGTCGAG	7
OS-DEC-004	GGCGCCGCCC	6
OS-DEC-005	GGCGCCGCCC	6
OS-DEC-006	CGAGGTCGAG	6
OS-DEC-007	CGAGCAGGCC	6
OS-DEC-008	CGACGCGGGC	6
OS-DEC-009	CCTGGCCGCG	6
OS-DEC-010	CCTGCGCGGC	6
OS-DEC-011	ACGGCCGCGG	6
OS-DEC-012	CGAGGACGTC	5'

The specificity of these octamers and decamers toward bacterial protein coding sequences was confirmed by frequency analysis in mammalian

5   DNA. The frequency in human DNA for each octamer was at least ten-fold less than in bacterial DNA, which was used as one criterion for selecting the octamers and decamers from the entire set generated. Additionally, a randomized search against known consensus sequences revealed no matches with most oligonucleotides generated. This confirms that these

10   oligonucleotides are indeed novel, unique, and useful for specific universal cloning of bacterial DNA present in a mixture. Furthermore, both the presence and high-level frequency of several of these octamers were confirmed within several desired cloning sequences (e.g. *S. clavuligerus ipns* gene).

15   Using this method, PCR with the octamer set has been clearly achieved using *E. coli* HB101 genomic DNA (gDNA) as template (Figures 9 and 10). When used as solitary oligonucleotides in PCR reactions, the amplification products were observed in the size range of 0.2 - 3 kb, consistent with that predicted from the calculated frequency of the octamers.

20   This demonstrates the utility of this octamer set for genotyping, in addition to cloning via amplification against a specific primer. A similar result has been demonstrated with *S. clavuligerus* gDNA used as template. Additionally, the ability to use these octamers as pair-wise PCR primers was demonstrated by amplifying a product using ACVS-04, a proprietary degenerate primer for the

25   *pcb* A gene (Figure 11).

The present invention is different from random priming and arbitrary priming in the following ways. Random priming is not specific for any type of DNA. Conversely, random primers are generally kingdom specific, as

opposed to RAPD, (random amplified polymorphic DNA method) which is a DNA polymorphism analysis system based on the amplification of random DNA segments with single primers of arbitrary nucleotide sequence. Instead, the present invention uses primers specifically designed from thorough

5   analysis of DNA databases, and the resulting oligonucleotides are universal for genomes included in the database. For example, in a method for RAPD PCR differentiation of Streptomyces species, none of the twelve 10-mer oligonucleotides matched the sequence of the over 65,000 oligonucleotides generated by the method for bacterial DNA amplification.

10   The use of the present invention also has a distinct advantage when the desired target sequence is derived from DNA of a mixed source, for example total purified DNA from soil. This population of total DNA will contain bacterial as well as fungal, plant, and potentially a host of many other contaminating DNAs, making it difficult to amplify specifically a product from a

15   single group of the constituent DNA, such as that of a desired bacterial gene. However, a universal primer set constructed as described in this invention allows for universal priming of a specific subset of the total DNA population, only bacterial DNA in this example. For example, a specific bacterial gene can be amplified from a mixture of bacterial and mammalian DNAs using a

20   single gene specific primer in conjunction with a universal library of oligonucleotides constructed as described in the present invention.

            Another example of the utility of the present invention is demonstrated by using it to amplify against a specific primer in order to clone the region of gDNA flanking the specific primer annealing site. Streptomyces clavuligerus

25   gDNA was used as template with specific ACVS, IPNS, and other specific primers to demonstrate the technique with high GC containing DNA.

            The combined results from all methods described in the present invention for the direct cloning of unique target genes from marine and

terrestrial microbial genomes is listed in Table IV. In summary, this collection contains 52 novel genes with homologies with the prototype gene or consensus sequence ranging between 35-90%. This includes a total of 10 classes of target genes, each gene within a class confirmed by

5   sequencing.(See Table IV).

Other adaptations of this invention center around the optimization and refinement of environmental DNA isolation and purification; PCR conditions and additives including DMSO, formamide, and others, use of neutral base substitutions and tails incorporated into the primers (such as d-azaGTP in

10   place of dGTP and inosine tails of 2-6 bases), and specific temperature cycling protocols; the construction and use of degenerate primers based on calculated universal primers, including the use of inosine in primers to increase length and annealing temperature; and the construction and use of labeled primers, such as biotinylation.

15   Cloned target genes representing the biosynthetic pathways or, in general, any flanking sequence, can be affinity purified from a diverse mixture of DNA, such as environmental DNA or total genomic library DNA. This includes both circular and linear DNA. Subsequently, the entire captured fragment containing the target gene/pathway is cloned and propagated in a

20   variety of expression/cloning host organisms and assayed for bioactivity based on the compound class of probe gene chosen. The method is based on RecA mediated homologous recombination and affinity chromatography.

Generally, the method consists of the following steps: i) biotinylation and affinity purifying the cloned probe gene; ii) reacting the biotinylated probe

25   with diverse, mixed DNA containing sequences complementary to the probe; iii) capturing the hybrid probe: complementary fragments on an avidin support; iv) eluting the captured fragments; v) and molecular and/or biological cloning of fragments and propagation in any suitable host, such as *E. coli* or

S. lividans.

Other uses of the novel cloned genes include hybridization screening, as exemplified abundantly by the data presented throughout this disclosure. For example, all probes/primers have been labeled with biotin and used successfully for the chemiluminescent discovery of novel target genes from southern blots of environmental DNA and genomic clones. Subsequent cloning and sequencing of these target genes was used to confirm that each probe bound specifically to its intended target. Thus, these probes are very useful (specific and sensitive) for the discovery and isolation of novel target genes, related gene clusters, and biosynthetic pathways.

The use of the DHFR2 oligos is especially promising for the discovery of novel folate antimetabolites, and their coding genes and gene products (biosynthetic enzymes). This approach is the only known source for the DHFR2 genes from which the oligos were generated as TMP resistant clinical isolates. TMP is a synthesized antibiotic and thus a search for a natural producer using genetic determinants for clinical resistance is quite novel (?). The DHFR2 oligo targets a unique form of DHFR protein that it unrelated to the chromosomal or other mutant forms that confer clinical resistance to TMP. Thus, the origin of this gene and protein have not been determined. DHFR2 can originate from a TMP-like biosynthetic pathway, conferring self-resistance to the producer. Following this model, the DHFR2 gene should be clustered within the entire TMP-like pathway. Thus, detection of the DHFR2 gene also provides the entire pathway within the regions directly flanking the gene. The results clearly demonstrate the utility of the method of the present invention have demonstrated the presence and possible origin of this unique gene in several environmental bacterial isolates, as judged by both colony hybridization probing, PCR, and sequence analysis of the gene.

ACVS04 (degenerate) and ACVS05 primers were used to PCR clone and sequence an approximately 500 base pair product from S. clavuligerus genomic DNA. This PCR was designed to generate 400 bp of known S. clavuligerus ACVS and 100 bp of new sequence of this gene. This strategy allows for assessing the accuracy of the sequence by comparison to a known sequence as well as generate new sequence. This confirmation allows for the routine use of the primers for generating new sequence directly from degenerate PCR products, a much more rapid approach than conventionally used.

DHFR2 has been used in the successful discovery and sequencing of several new DHFR genes. These genes confer resistance to TMP and other folate antimetabolites in WT as well as clinical isolates. Additionally, many of the WT strains produce novel folate antimetabolites.

The above discussion provides a factual basis for the use of the methods and probed of the present invention. The methods used with and the utility of the present invention can be shown by the following non-limiting examples and accompanying figures.

EXAMPLES

GENERAL METHODS:

General methods in molecular biology: Standard molecular biology techniques known in the art and not specifically described were generally followed as in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York (1989), and in Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Maryland (1989) and in Perbal, *A Practical Guide to Molecular Cloning*, John Wiley & Sons, New York (1988), and in Watson et al., *Recombinant DNA*,

Scientific American Books, New York and in Birren et al (eds) *Genome Analysis: A Laboratory Manual Series, Vols. 1-4* Cold Spring Harbor Laboratory Press, New York (1998) and methodology as set forth in United States patents 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057 and incorporated herein by reference. Polymerase chain reaction (PCR) was carried out generally as in *PCR Protocols: A Guide To Methods And Applications*, Academic Press, San Diego, CA (1990). In-situ (In-cell) PCR in combination with Flow Cytometry can be used for detection of cells containing specific DNA and mRNA sequences (Testoni et al, 1996, Blood 87:3822.)

Recombinant Protein Purification

Marshak et al, "Strategies for Protein Purification and Characterization. A laboratory course manual." CSHL Press, 1996.

EXAMPLE 1:

1. Preparation biotinylated sequence-specific probes.

40ul (100pmol/ul) DNA primers StsC03, StsC04, IPNS05, IPNS06 were labeled with S-S photobiotin (Vector Inc., CA) according to the manufacturer's instructions. These primers were designed for the amplification of the stsC and ipns genes of *S. griseus* and *S. clavuligerus*, respectively. After n-butanol concentration and EtOH precipitation, the primers were diluted in 200ul water (20pmol/ul). PCR reactions were carried out from pT7 blue3 (Novagene) plasmids containing the previously cloned probe sequences pT7stsC (*stsC* from *S. griseus*) and pT7ipns (*ipns* from *S. clavuligerus*) as templates.

Reaction mixtures contained the following: 2ul primer 1 (20pmol/ul) (StsC03 or IPNS05), 2ul primer2 (20pmol/ul) (StSC04 or IPNS06), 2ul template DNA (10-100ng/ul) (pT7blue3/StsC/S.gr or pT7blue/IPNS/S.cl), 2ul

buffer, 2ul dNTP mix (2mM), 10ul water, and 0.2ul (0.2u) Taql. Cycling was conducted as follows: five minutes at 95°C, 30 seconds at 98°C, 30 seconds 52°C, one minute 70°C, repeated 34 times. Finally, the reaction was heated to 70°C for ten minutes followed by holding at 4°C until analysis of products was performed.

10 ug of the mixture PCR product was purified on Avidin DLA beads (Vector) to separate biotinylated and non-biotinylated probes. The yield of biotinylated probe was 4-5% (0.4-0.5 ug, Figures 12 and 13). The biotinylated fraction of the probe was used for RecA capturing and non-biotinylated probe was used for alk-direct labeling (Amersham) in hybridization screening.

2. RecA capturing of specific target gene containing cosmids from pFD666/S.gr/library.

5ul (0.1ug) of the biotinylated probe was denatured by incubating for ten minutes at 99°C mix with 50ul RecA buffer (25mM TrisAc, pH 7.5, 10mM MgOAc, 2mM CoCl$_2$, 1mM ATP, 2mM ATPγS, 5ul (2ug/ul) RecA (NEB). The mixture was then incubated at 37°C for 30 minutes. 2.5ul (2ug/ul) of the CsCl purified cosmid DNA was then added and incubated for an additional hour at 37°C. 5ul (50ng/ul) of lambda HindIII digested DNA was added to the mixture to remove excess RecA, incubated for ten minutes at 37°C followed by the addition of 2ug/ul Proteinase K and SDS 0.2%. Enzymatic digestion was carried out for 30 minutes at 37°C and then reaction was stopped by adding PMSF (100mM) to a final concentration of 3mM.

The captured DNA was separated on Avidin DLA beads (20ul) beads prepared according to the manufacturer's instructions. At the final step captured DNA was eluted with 100-200 ul (0.1M NaOH, 1mM EDTA), EtOH precipitated, and dissolved in 20ul water. DNA was electrotransformed into *E. coli* XL1 (Stratagene). Positive clones were detected by colony hybridization

with alk-direct StsC   probes (Figure 14). DNA from positive clones were purified and verified by dot-blot, southern hybridization, PCR and bacterial growth on LB agar with streptomycin (20ug/ul) plates.

3. Direct cloning of RecA captured DNA fragments from S. griseus chromosomal DNA.

RecA capture was carried out as described above, but instead of cosmid DNA, 5ul (1ug/ul) of chromosomal DNA from S. griseus digested with Mbol/Sau3Al and CIP was used. After RecA capturing and binding to the Avidin DLA beads, DNA was eluted with 200ul 2.5mM biotin and directly ligated into the pSCOS1 cosmid vector (Strategene). After packaging into lambda extracts clones were plated on LB agar with Amp (50ug/ml) and Km (25ug/ml). Positive clones were detected by colony hybridization with alk-direct StsC probes (Figure 14). DNA from positive clones was purified and verified by dot-blot, southern hybridization, PCR and bacterial growth on LB agar with streptomycin (20ug/ul) plates. Positive clones most often contain related pathway genes, as confirmed by additional hybridization with related gene probes, such as strD, strB, and stsC (Figure 15), and PCR (Figure 16). Additionally, heterologous expression of these genes is often observed, as judged by antibiotic resistance (Figure 17) and HPLC chromatographic profiling of cell extracts and fermentation broths, further demonstrating the utility of this invention for expression cloning screening.

Results from both examples clearly demonstrate the advantages of targeted cloning in providing highly enriched libraries of specific genes and associated sequences, including biosynthetic pathways. Library enrichment of several hundred fold for specific genes and related biosynthetic pathways it has been demonstrated by use of the present invention (Figure 14, 18).

Throughout this application, various publications, including United States patents, are referenced by author and year and patents by number.

Full citations for the publications are listed below. The disclosures of these publications and patents in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

5   The invention has been described in an illustrative manner, and it is to be understood that the terminology which has been used is intended to be in the nature of words of description rather than of limitation.

Obviously, many modifications and variations of the present invention are possible in light of the above teachings. It is, therefore, to be understood
10   that within the scope of the described invention, the invention can be practiced otherwise than as specifically described.

TABLE I

List of example degenerate PCR primers of target gene cloning

Gene	Function	Pathway	Name	Sequence (5' to 3')
aac6'	Aminoglycoside 6'-N-acetyltransferase	aminoglycoside	AAC6001	CGATGCTSTAYGARTGGCTA
			AAC6002	TGGCGYGTYTGVACCATGTA
			AAC6003	CCGACRCTYGCKGACGTACA
			AAC6004	CATCBGGSGTSGTTACGGTA
acvs	alpha-aminoadipyl-L-cysteinyl-D-valine synthetase	beta-lactam	ACVS3	TGCTSGTSGGSGARGAGCTGA
			ACVS4	TCSACYTTGCCRTTGACRTTGA
			ACVS5	TTCACCGARRCSGCGTTCGTCA
			ACVS6	CGCCGGSACCATSAYCCGGATSA
dhfr2	trimethoprim resistance		DHFR2-1	GATCGCGTGCGCAAGAAAT
			DHFR2-2	CCACSTTTGGYHTRGGRGATCG
			DHFR2-3	AYRCGTTCAAGYGCMGCMACAG
			DHFR2-4	GATAAATYTGYACTGARCCK
ipns	isopenicillin N synthase	beta-lactam	IPNS3	TTCKSCGAGGACCACCCGMWGAT
			IPNS4	GGAAGTAGTCGTKSGTGAKGT
			IPNS5	GATGCACGAGGTBAACVTCT
			IPNS6	TCTGGWASAGSACGGTGATCA
NIM	Nitroimidazole resistance		NI01	ATGCSRCGYAARCGGCARTTGT
			NI02	CGATRGCYTCYTTGCCYGTCAT
			NI03	CGTTGGCYCTTCATGGSGATGA
			NI04	CCTTTGGCTATYTCRGCYTCCAT
PKS	Polyketide synthase		PKS1	GAGTTCGACGCSGVSTTCTT
			PKS2	GGTGTGNCCGATGTTGGACTT
			PKS3	GCARCGVCTCCTGCTSGAAA
			PKS4	TTGCTSGCRCCGTCCTGGTT

Gene	Function		Name	Sequence
strB1	Amidinotrasferase l	aminoglycoside	StrB1001	AGCGTSTTCGCSGTGGAGTA
			StrB1002	GTGCTTCTCSAGMAGCTTGA
			StrB1003	TSAAGGAGACCGARGASGAA
			StrB1004	CGSACCACSAGCAGGTTCA
strD	dTDP-glucose synthase	aminoglycoside	StrD1	CTTCTAYGSVCTGGAGGCCA
			StrD2	GAGGRGATCTGSSCCDTGCT
			StrD3	GTGGGMGACGGYTCSAARTT
			StrD4	TYGCGCAGCACGATGGAGAA
strE	dTDP-glucose dehydratase	aminoglycoside	StrE1	CACTACGTCCRSACCCTCCT
			StrE2	GCCAKBCCGTCGSYCAGGT
			StrE3	CSGYSTTCGTGCGGACCAA
			StrE4	CGGTCSKSGACGTRCTCCA
stsA	L-alanine:N-amidino-3-keto-scyllo-inosamine aminotransferase	aminoglycoside	StsA1	ATCGTSCCYGGRTASATGTT
			StsA2	GATGGARCGSCCSAGSACA
			StsA3	CTTSTCCCTSAAYCASTAYAAGA
			StsA4	GCGTTCCRGGYTCRAAGGAA
stsC	pyridoxal phosphate dependent amidotransferase	aminoglycoside	StsC1	GCMTSCCSSTSATCGAGGACT
			StsC2	GTCGAACCGGGSMSGGGTCCA
			StsC3	CGGCGCRYSGGRGTSTTCA
			StsC4	GTWCAGCGGGTTGKCGTTCA
Taxol	taxadiene synthase		Taxol01	ATGATGTGGGTYTGSTCSAGA
			Taxol02	TTTYTCRICSCCYGTRTTCAT
			Taxol03	TCCYGGYCCKGTSGTMATGAT
			Taxol04	ACCYTCSAASGCRTTYAACAT

TABLE IV

Unique Cloned Genes Used as Probes

Gene	Function	Clone Name or Accession No.	% similarity with prototype	Sequence
aac6'	Aminoglycoside 6'-N-acetyltransferase	AF034958, U59183, X60321, L25666, S49888, S45954, U90945		
		LDSAA3-33	50.80%	ccgacactgcggacgtacag
		LDSAA3-51	49.50%	ccgacgcttgctgacgtacagg
		LDSAA4-56	36.10%	catctggggtggttacggtacag
		LDSAA4-57	48%	ccgacgctcgcggacgtacatg
		SCDSAA2-49	40.60%	catcggggtggttacggtatlc
		SCDSAA2-52	36.20%	catcggggtggttacggtataa
		SCDSAA3-85	48.60%	ccgacgctcgcggacgtacatc
dhfr2	Trimethoprim resistance	K02118, X04128, A12434		
		pASDMN1	89.80%	gatcgcgtgccaagaaatctg
		pASDMN2	87.10%	gataaatatgcactgagccggg
		pASDMN5	89%	gataaatctgtactgagcctgga
NIM	Nitroimidazole resistance	X76948, X71443, X71444, X76948		
		LDSN1-12	45.50%	ccctttggctatttcggcttccalctc
		SCDSN1-22	49.10%	ccctttggctatttcggcttccatcg
		SCDSN1-28	48.10%	ccctttggctatttcggcttccatgc
		LMSN1-65	47.30%	cgatggcctcttgcctgtcattt
		LMSN1-86	48.90%	cgttggccctcatggcgatgat
		SYDSN1-83	47.60%	cgttggctcttcatggggatgatg
		SYDSN1-88	48.10%	cgttggctcttcatggggatgatc
PKS	Polyketide synthase	AF007101, AB032367, M63677, AF016585, AF079138, U24241		
		LMSK1	42.20%	ttgctggcgccgtcctggttggtt
		LMSK25	49.10%	ttgctcgcgccgtcctggtttgcg
strB1	Amidinotrasferase I	X78973, Y00459, AJ006985, X78972		
		SCDSB31	47.60%	aaggagaccgaagaggaaca
		SCDSB32	39.80%	tgaaggagaccgcaggagga
strD	dTDP-glucose synthase	AJ007932, AJ006985, Y00459, AF055579		
		LDSD2-16		ttgcgcacgacgatggagaact
		SCDSD1-30	49%	caaagcggaagcgatgcggat
		SCDSD2-40	45.30%	tgcgcacgacgatggagaaag
		SCDSD2-44	48.90%	tcgcgcacgacgatggagaat

acatgaacctggccgctcggttggggtltagaagaagataccggtctagagcacttgaagtaggaggaggccggaagggatgggggaaaacgctctcacaggtgacaagggaacgaggcagggtcttatctтaaggctgatctccgagaa
cggaatgcagaactтccacgcgggggcattaaaacgttcacgaaacggcagtagtttgcggtgtcaggccgatttccggcaccatcaccagcgcctgttгccctgagcgcagcacgtttccagtacgcgtgagataaacctccgtttac
gcaactтcacgcccatctgccgagtтcaacatcttgccgaccccagaagccgcacgcтagttгttcacctccggcgttcacctccggtcggagttcggcccgacgtgattgctcggatggaaagg
acaaaccttattтcaagatтaagaagataagcgcaaggtcgagaggtgaataatgccтtgctgctcatтggtggcgcgacgcaattgctcatagcactcacgtgttaatcactccggcccaa
gaggcgatatcattттcacaggaattacgcacccaaagactcaatcagattтgcgtccaacaagaccggcattg
actgcggctтcctttтctttтctттcttтcttтcttgttacacgtgtaaacaacagaagactgcttagcgcaatacttgccgacaa
ttcaagcgtcatcgctcggcatgttgatcttgatctgtgataccccgcatacgcaaggggtggggcttgcgcgcggagcgttcgccgccatcatgcggtgacgacggtcgccgaagcgtaccgg

cgccgcctggcagggtgcgagttgtcggctggtactgcacagatctgaccccctgaaggctatgccgtgagtcggatcgtgtccagtacagatttatc
gtgagactcggactcgacggcatagcctcagggtcagttgtcagtaccagccgacaactcgaccctgccaggcggccagattтcttgctgcgcacgcgatc
tgcgactcggactgcacggcatagcctcagggttcagtttгtgtcagtaccagccgacaactcgaccctgccaggcggccggccagattгcttgcgcacgcgatc

gacgacagtgaagttctccctgagaagaagtcgaggtaccccctgttctgaagaaaтgtccтtгgtattcgaccgtggaccgcctтгgtatcgagcgctggaaatccgagggtatgggggcgaggtcggcataggctggaacgcatt
tggcctccttgcctgtcatccactтctcaacagagatattтgagaaatcagaaattтctgttтttaaaggaggatgtcggctgcgggaaccgatcatctgtagctgttтctttataaattcgacgcтtgttтctтtctгgctттtттctaaag
cgtcatgacaccctcctggttgttgctacaaтtттctттatcacctтgccctgtтctтcttcacacgtcaacggacttactaccatcggtaaatgccgcggcgtatcatатtтccgcctcatттctтtccacccтtcattтcctcctгatcctгatcttгacttactcaggtaac
tcgatcactaccaccgggcgtcgcagtcgtattgccagcgccctgtattgccagtcgcggccctgtcgccgtctcgcgtcctтgtтctcaatacaaaac
gtcagcaccacgcctgtcgcggcggctcaaaagatagctgtggccgagcagatgacgggaaaacatgctggccgatctgtcgcaaccggcgatccggagатctggacggccaagaттggtgacggccgt
gtcagcaccacgcctgtcgcggcggctcaaaagatagctgtggccgagcagatgacgggaaaacatgctggccgatctgtcgcaaccggcgatccggagatтggtgacggccgt
tgccggctggagggggtcgcgcgcacggaggaaagaataaggctcgtaggtcatggccgcgtcgttгctggccgcgatgaaggctccgctccatgttгacgacggtcagcggctcctтcacggagagaccgттcttctт

gcacatgccattacccatтacgatggtatcatcaccgcgatagccgaaaтtгcaccgcctctcgctgctgttтccctcataagagacctcataagcgaatттactacgctcggacaaaacccattcacagccaatacccgactgactc
agacgagggcaagtgcagtaccagtгggagagatggtaccgcccatтcagcgagagagtagcgttaagcaacaatcgcggtaagaaaccatcgacattгgcgagaaccagcacaacacggtтccgggcgct
agacgagggcaagtgcagtaccagtгggagagatggtaccgcccatтcagcgagaagtagcgттaagcaacatcgcggtaagaaaccatcgacattгgcgagaaccagcacaacacggtтccgggcgct

tgggctggaagatgaaaagccaaaaggaacctттacgtatgggcgtgtagaatтccgagaaacgtттgagaacatctcaccaatтctgattacttgctggagcatgctcatgtcgttгtcacacggcggaaaatattcggaagca
acacgaaaataagaacgatgtgctgcgcaggaagatcagcaggggcattaacagggaatтaaaatтaagacaccagaggaaatataagacacaggaaataatgatcaaacattccтaactataggagcagagcgaggcattaatcatт

galttгcgtgtcctcgatccgatcggcgtтctcggagacggacgggacggtgatgtaggcgtgtтggaccacgtgcgttгgaccacgatgtgggctgcaacactgacgcтtgcgcgcctтgacccggaccc
gttгgcgggcaaaagtcgcggccttagcggcgcaacaggcgtctgatgaaaatgatcgaggatggtggtcggataggggctcgagggtgcgtactggcggttгgatggcgtacctataactттcccgaggaaatcgcctggac
ccggctatatggacgagagattctтcctatatgccgagaagttгacgtaaggtgcagccgттacgtaacgcgggcgaattagcgatcтtтggagacatcaacatcacttгcacctтcaggttгgagacaccagaaccacggagaccgcctтtgactcagccgat
tgggtatcgtgaaggtaattгgagaggagagatataaagtaccagcgttгcccccgaaттgttгctgacctgcatacagggccttгacgtcatacagggccttгatcggcgtgagtcagagagcctтatcggcgtgagttгatcaaagggc

gcattgggaccagcaggatctgtcgcgccctgcgggagctgggttgagttggcccaggcgaccgtgagtcggagacctgcgagctcggggtcctaaaggtcggtaaccgctatctcggc

gtaaccacgcccgatgatcacgaattcggatccgatacgttccgcagctgggtaccgagcttccctataagtgagtcgtataga

caggcggcccgccggagagctgttcagcgacatgaactcacctcaaaacacgcctt'gctggcggccggtgcacgacgccac

tggtaaccgatggtccctgaagatgtccagccctaccataccatcaccaaagatattggtgtgaggcactgatgtcaccggacacacggaaaagaccatcattgctccggtaa

aataccgtaaccaccccgatgatcacgaattcggatccgatacgttaccgctgcgagcatcgcggtggtacgagctttctataagtgagtcgtatagagg

tcggaaccaggtaggtgggttccggggaggtggcctcggtcataggacaaacgtccgaggatcattcacgtcgcccaatgggcggccggttggggccgt

ataccagaatagcaaccaaaggcagcaagcagtacaacacctgccgtttggccgccatatcgtgaattcgtcgacaagcttcttgacctaggcta

tatatcaccaggtacgtctattcatcgccatgaagggcccacgatctgaattcgtcgacaaggcttctcgagcctagggcta

atggaggcgtaaagcgcaattgttcgacaccgagatgacggcaaggaggccatcatcgccatgaagagcacgatcacgaattctggatcg

attgaggccgaaalagccaaaggatcgaattcgtcgacaaggcttctcgaggctaggccagctcgaggggcccagctcgcggcgcacaattcactgc

cagcatccaggagaggcgaaataggcgacgtgccgggcgcggaggccgccacctgctggccctgatgtccttgatgagggccgagatagccaaagatcgaattcgtcgacaag

atccctttagaagacacagaataatgcaaatcacttgtagctacgtttcaagatatacacattattgtctcaattaattatttttattaggatagataggttggaccat

agttttgatggtgtaacgttagatgcggcgatcagttcgttcacctcctgccaggagaacgaacaatccaccgccgtcacggcgcctgcttaaggcttgcgct

cggaaaaggcatgtcagaaatcgatggtgtgaagcaggaggatctgcgggaatttgtcatgcggattcaaagctgaacctgctcgtggtgcg

accaactatttcaacaacatcagaattgaataaagaaaaatatttgaaaatgctgtcattttgacagctttataaactactgaactgctagtggtgc

ggccgatgatctgctgtctcgtgtcgaccgtgccctgacccggctccgacggtctctcaaggaccgaccggtggacctccaagcatgtcgggtaacgttgccggtgtccttccaggtagccgagagaaaggtccgttggaggcga

gggatcatcgtaattgggtacaagttccaggaaacttgaccagagttctggtggcggaccaggttctaggacgcggctccatcgccgataggtggagggcgtcggatggcacaacggccgaaggtcag

aaggctactacggcctgatgaccgtaaaggcctcaatcatgtcaggtcagaaacaatcgggcacgctggtactat

gtctccgggtgggtgctcaccccgggggtgagtgatggtgatggtgcgccaaagagtcgggctaattggggggccagcgttacggtgaacgggctgcgaggcac

acgtacacggctggctgtgggtcaatacagcacactgtgaatggctgtgggtgaaatctatcaggctcggccggccgcacagagaccgcgctcttggtgaattgccggtgataaaaa

Gene	Description	ID	%	Sequence
strE	dTDP-glucose dehydratase	SCDSD3-57	48.70%	cgtcgttgagcggacatgcgct
		SCDSD3-62	48%	gcttgcgggtggagatgatcttg
		SYDSD1-19	45.80%	cgcagcacgatggagaagtgt
		SYDSD3-52	48.30%	cgcgcacgacgatggagaatc
		AF055579, AJ006985, AJ007932, X62567		
		SCDSE1-79	51.80%	tcggtcgggacgtgctccacac
		LMSE1-57	48.40%	ccgcgttcgtgcggaccaactg
		LMSE1-63	49%	catggataacgcctggcaggg
		LMSE1-68	47.40%	cactacgtccgcgaccctcctg
		LMSE2-84	49.80%	cggtgttcgtgcggaccaaaag
		SYDSE1-54	35.80%	cggtccgggacgtgctccagac
		SYDSE1-55	37.70%	cggtcgtgacgtgctccaggc
		SYDSE1-61	49.10%	ccagaattcgtgatcggtgttcgt
		SYDSE2-66	35.30%	cggtcggggacgtgctccagg
stsA	L-alanine:N-amidino-3-keto-scyllo-inosamine aminotransferase	Y08763		
		LDSA3		
		LMSA1-6	47.70%	gcgttccgggctcgaaggaaa
		LMSA1-9	47.20%	gcgttccgggttcgaaggaagc
		LMSA1-17	47.40%	gcgttccgggttcgaaggaagg
		LMSA1-26	48.80%	gcgttccgggctcgaaggaata
		SCDSA2-18	52.80%	ggcggttccgggatcgaagga
		SCDSA2-20	44.00%	cgtagagatggggtctctccatg
			48%	cagcgcggcagtgggtgggtta
stsC	Pyridoxal phosphate dependent amidotransferase	Y08763		
		LMSC1-29	48.60%	gttcagcgggttggcgttcagaa
		SYDSC1-1	50.20%	gttcagcgggttgtcgttcatggc
		SYDSC3-22	46.10%	gttcagcgggttggcgttcaggg
Taxol	Taxadiene synthase	U48793		
		LDST1-81	49.50%	tcccggtccggtggtcatgatcct
		SCDS1-33	40.80%	accgtcgaaggcgtttaacat
		SCDS2-24	47.10%	tcctgccggtcgtcatgatgt
		SCDS2-25	50.20%	accctcgaaggcgttcaacatc
		SCDS2-42	49%	tcctggtccggtcgtaatgatcc

aaaggcagtgaatattcgcgctggttgagagaaagcgatccggtagcggaactggcattgctgctgctacgagctgcggctggtggcaaaatgctggcacatgctgtgaaatgctggcgaacaggatgcggcgcctcggcaattgctgatccgatgtgattgtcctgggggcgcgggatgag
gcctcggggatcgttcatcgccaggatcaccggatcggttgacgtcggtcattccaggcttcggcgacgtcggtcgctctacaaatgcctttacgggcatcgatcaggaggaaacctgccattggaagcaccggtatattcgatgtgacctgggggtggaaagtagatatg
ggaacgtctgctctacaaatgcctttacgggcatcgatcaggaggaaacctgccattggaagcaccggtatattcgatgtgacctggtgtctcaataatgtatttacggtggggtggaacggcaacggcagacagcaagcgcacctcgcggagcagacggcgacactggccgacgacactggttcatcagaacgctttggcgg
gatcgggtccgccttcaacgatctgttgattggcgacacagtcgaaaggtgggaacggcaatgacaccttccgcgcacgcaagcagtgcaatgacacctgtcatcagaacgctttggcggatcccagcatatcgatgcgccggcacaacagcagcgaaatcacggctgcagaccgcca

gcgagcaacaactaccaggacaagcaccaggccctgtccgctatgcaaacgtgatgacgtgcagccgcaccaagtgccctgccgcccgggccgcggctacaacagcagcgaaatcacggctgcagaccgcca
gcgatatagaaacgggccaggcagccgttggctgcgaaaatagggcctgatctccaggtgcatcggcgggctgatgaggctgaagttggcaggtgagtggccggctgagagttggcaggtgggttgacggtc
aacacctactgtccaacgtcggtctgttccgaaggtggtgtcaatcagtggtggatcagatggtcacaagtccttcagctgggtcatccagctggtcacaagccgtgaaaagccctgctac
aagtgcagcaattctttcagcccgccgccagcgacatttctttgcgcgcgatacggcgctcatacagagccattggcgtgttcatctgtcgtttgtttcatcgtgggtttaaccgctttcagtttcttac
gaacgaatttcagacatcagcacccaactgaacgcctttccgcgctgtgaagttgcttcagcgacggcgccgagcgcgtgatcagcggcaatcagcatcagctgcagtcgcaacccgcggaatcacggcggcaaataggcgctgatccagaacctgga
gctgcgcgaccgaatatgtgaagaccgaaaagaagcggctcgtcccgcagacaaagcgcgatcgtcccgcctcctggagagcttcacggcgacgtcccggatcggaaatcacggctgcggaatcacgtgatcgtccgagagcttcacgagacaatcagcgcatcaggctggctggtgctaatcgctgatggtcc
gacctcgtcaggctgaggctgatttcatcgaccacctgacctcgtcaggctgaggctgatttcatcgaccacctgatcgtcgatcgccagcagcgacggcgcgtcatcacctgtgtcgaactggccagagtgtccgc
cggaccaatccgctgatcaactcgccgggcctgctgataagatcggatcaggtgctcaaacccttaggacctgtcgtcagagcgaaggggaggggaggagggactgtattccaccatctgtcgaactggccagagtgtccgc
aaatgaactgatcctctccgcgggcctgcgatagtagccggatgaagcgcacgctgaatagatggtgaaatagatggtgaaaagccgccgctcgggcctgtcagtgccgaggca

cgttcagaaggtcagctatatcgccggcgattctttgcttcgtacctgcgcgacgaccgcggcaccgaagtaagagatgtacgagaggcgcagagtctgggcgtcgcgtcgcgtgttcgcggtctcgggcgcgcgtgtcgggtcggtgttgaaggacggaa
aacttccagcaggggaacgcctcatcctggcatcgttgctgatcgttcaacagatcgttgctaatttccaacagaatgcggtccaccatcagaatgcggtgatcgccatccgcggcatgtgaattgatgacgccaccatcaaacttcat
ctggacttaatgagcaaggagcgaggtaatgcgggcagcgccctcttgtctcctgtctcctgtctcctcctgatcgtctcgtaaaacgatactgggaaagccggccctctcttgagtggcggtgttgagaacgccgggttctggcaactcac
ctgtctcatgaacaggatatgtcgtcttgcatcatgatcggcatgattctgcatgattctgcatcgcatcgatctgctgatctgggcgcacgtcgttgtggagttaaccagtcagactgtatgacattat
accgttcagaaggtcagctatatcgccggcgatatatcggccgtatatcggccgttctgtacctcgtccggtgtctcccaaagtagtcagggtttaaagtgctgggtattaaagtgctggggtattaaagtcatgatcca
tgccaggctgtatcgtgactcctgggctcaagtgatcctcgctgctcctgtcctcccaaagtagttattcgcttatacgtcacttgatctgagttacttcttctgtgttcttggtgctgcgttgatctcttagttcttagttagttctagttgatctggctctctgagttgatctctgacg
ttgctcgttagctgcgttactggttgagccggggttgctttctactcgtaccgtcgtcacgtcacggacgtctttcttctcacgtctttcttctcacgtctttcttctcacgtctttggtttctacggttcttgacctgtcacg

gcagcgtggctgggtgccggatggtgccggatcgtcgacgcgcggccgatggtgctggtattccgcaagcgacgtgcgcgatcttattgtgacaagcgt
cagaccagcagcggtatcgtctccaggtcttctgcgatgtggcacacgcgggacatgccgtctgctgcgcggtaacagacaggcagctgatcggccgctctcgc
attggtgcattgcctgcctccggtgcctgccggtggaacctgaaatcccggtgacttttggcatggcatcagcgtgggcatggacatggcagtcaatcggtattatgatgttgctgcctgtaccc

tcgcctctgctcacgaaagatgctgtccgcccatcggaagaactcactatttcggtcgttgtgtgttgtttggatccccgggtcccgggagcccgggagcccgagccgctcattcgagaggtgggacgactttgagagaggaagaaacgctgg
gccggtgatgagttcacggaagttcacgggaagtcgactacaggaatgggtggatcaacctgaaagtagaatgggcgattcactactgaaacagaagaaacggtacgattcatacttgaaacagaagaaacggatcttgg
tcacgttattatgtagtctgccggacaccttattcaggatgagtatcagcagagagtgaactacagatatctgttggactacagtctgctgatgctcgtctacgatctgcacatcagcagatgctcacgatctgcgcgaaaacaccccc
gccttcagcctcattctcagtagttaatgccatcggtatgggtgaatgcatgaaaatggcacggcggaatcagaatgggcggaatcaagagggagaaaatgcagcaatatggcggaacacagaatcagcaaaggagaa
gagctcgcagcaatttcagtactacggaactgaaactgcagcctcatctgaactgttatacctattatacctacctgcgcgacctattattatacctattacgtcatcgcagccttatcacctgcagaatttcagtgcttaatctggaat

caatgtagaccgttatatcaaacggttgggcagttgattaacaatttggtcttccgcgcggcagtgaacgccggtgcgtaggcgacgacgggtgaatcacgagttctg
gggctgaccaggcgatagccttttggcacttcaggtgggtctaggcggccgggccggtggccggccatcgcccatgatcaggatctgccatcgccagccaccggttgctcgt
gccacatcaatggtgataacctgtcacgttcagccacaaggccgtctgcagcaatgacacaggtctgtaaaatcaagtcctttgcgttg
gcgtcctcggccggcatcctcggctgcctggtcacgttcactgactgctccaacggacaacacagacacaggtgccgggggggaccgaactagagtgttctaatgcgagctagagccatgct

cgtggtcggcccggcggcgaggaaatctacaccgacgaatatgccggtgcgctgcagttccactggacacgggaggggcgcgaacgacgagcgcagggtcagcctggataccgctcgtccgcac
ggcgagcatttccattgatacagttgctctgtgagcaggctttccaggtgtctctcgtcaggttccgacgggtgatggagccagctgggaaggactggtgagccactctg
aggagcaactgtgatcaatgacatgcttggctgaccggtcaccctgctgagccggcggctacctgccaatcagccatcagcggagtgcacctgggagtcaaggcggccataggacaggcca
ggtttacctgcctcggcaaaacgtcgagcattcaggatcccacccttgaagggtcaaggttaagggcgcattgcagataatcgcttgagctctggtctgcgttttta
gtagagggcgtgctgcggtgtcggtgttatcaccagcagaagagacaaggtgagaaacaccatgaaactcagtcgtgctgctggggcagaagcagctggggctcggcctcgccgttcggcggcgcggtgccggtctc
gcatctccaattccgagatccgagctgggaagcaggtgcttccgcgattctggcggcgcacctctggcagccatcggcgagacgaggagctgccgcaccgcggagt
cacgaagacagccgttcgtcggacaacaggggccgtactctgtattcttcagaaggatctgggaagaatcaggtagcggcttggttctccgacctgtagtgatgcccagc
gctgtgatcagalcctccaggcttccaatcgggcgataaggccgatccagccgcggtcgcataagccccgaataagccaccggtcgcacacgagtgctggcattctccacgagtgtccacagagtgcggcttccacgagtgtcggcgcggcctggggcgttgacgcg
ggcggatgccggggcgccgcccgagccgaggccgaaatagccggtcgcacacgagccgaaatagccggtcgcatagcgccatctggtgcggtggtagtcggcggttgacgcg

acgagacggaacgttcacgtattcacaagctacacgttccctcgttgtttaccactacgatttaaagaccaacaagagcactcttgaagcaacccgaaggtcgacgcggatctacgaaatatgagaccagcctcgtcttctacaaca
tcacaggtgaggtttccaggtgggcatcatcatggtatcggatcgaccataaggccgtaatcaccaggttttggtcgggaactggtcggagaataaatcctctgcggttcttctcatcgccacgac
caagctggcagcacagtttatttcagagagatgaccgttcaaggtcatgttcacggccatcgtcgtcgccatggtctgtatattcgcgactcaggtctggggcttcagacta
agcgcgttaaatcttctggtgcgatgggatgtttgctggtgctgatcgcagcatcttctcaaacagttgccgaaggatttcttc
gggaagacgagacggaaaacgttcacgtattcacaaggctacacgggtcacgggtcctccggtgttaccactacgagttaaagacaccacaggagcactccttgga
aaagactggagtatttgtcaatgaacagttcaacatatgtatctttcaacaaaatgcagcgtggtttaaatcctaaagc
ctcatgccacggtgacaacgatgagttcccatacagatccagcttcctggcggggggtggagtgtggacaaaggggccttgatcgcaaatcctgcaccaccgtctct

gtctgtcatatacaggtaatacggcgctcgagaaaagctgactcacgggcacgacatttgataggcgcttaagctgctgccactgctgctggact
gaacatcgcgagcgatacgcccgtccatccgcgcacgcgaccccgcattgtccaggatgccctccgcctccggcctccgaagaacgagcgggccgt
ctacgtacggcaatcttggcttagcagtcagttgcagttggtgcatggccgtgtg

cgccgggtgatggaaggcacacagtgtcaacgcgagataccgattggtccatcgttcgtataggtcatgtgctcaactacat
atctggaattcgttcggacaaagctttcttcggagctggctagctagcttctagacaacaacgtgtggggggccgagctcccggccgcaacaattcacattgggccgtcgtttacaacgcttgttgtca
cataccatccgagcgagcgtgattataacaaacgtcttccgacaagcgagagcctcgcgctctggatagagatacatcgtcagattac
atgatgtttgagactactcttgctgccaggagactacatgccgaaagcagaagcggtacacatcaaaagagatacaacacaggcgggaacatgctgtgatg
aggctgtctgtaatttctttgcatctcgcttattcaggtgtgttgtgttcaggtgtgttgcaggggagc

caaagatggcacgcgcgtaccattgttcatcaccgcgcgcaaggataagctggacggaacagaatcccttacccctatgatacggcggatc

cgcgtccgcac
ctg
taggacaggcca

cgcgtgccggtcto

gcg

aaatatgagaccagcctcgtctctacaacagcaaagatggcacgcgcgtacoattgttcatcaccgcgcgcaaggatataagctggacggaacagaatccccttaccctatgatacgccggatc
cgac

gcttgttgtca

gctgtgatg

Substitute Sheet (Rule 26)

REFERENCES

Burke and Olson, "Preparation of Clone Libraries in Yeast Artificial-Chromosome Vectors" in Methods in Enzymology, Vol. 194, "Guide to Yeast Genetics and Molecular Biology", eds. C. Guthrie and G. Fink, Academic Press, Inc., Chap. 17, pp. 251-270 (1991).

Capecchi, "Altering the genome by homologous recombination" Science 244:1288-1292 (1989).

Davies et al., "Targeted alterations in yeast artificial chromosomes for inter-species gene transfer", Nucleic Acids Research, Vol. 20, No. 11, pp. 2693-2698 (1992).

Dickinson et al., "High frequency gene targeting using insertional vectors", Human Molecular Genetics, Vol. 2, No. 8, pp. 1299-1302 (1993).

Duff and Lincoln, "Insertion of a pathogenic mutation into a yeast artificial chromosome containing the human APP gene and expression in ES cells", Research Advances in Alzheimer's Disease and Related Disorders, 1995.

Huxley et al., "The human HPRT gene on a yeast artificial chromosome is functional when transferred to mouse cells by cell fusion", Genomics, 9:742-750 (1991).

Jakobovits et al., "Germ-line transmission and expression of a human-derived yeast artificial chromosome", Nature, Vol. 362, pp. 255-261 (1993).

Lamb et al., "Introduction and expression of the 400 kilobase *precursor amyloid protein* gene in transgenic mice", Nature Genetics, Vol. 5, pp. 22-29 (1993).

Pearson and Choi, *Expression of the human b-amyloid precursor protein gene from a yeast artificial chromosome in transgenic mice.* Proc. Natl. Acad. Sci. USA, 1993. **90**:10578-82.

Rothstein, "Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast" in <u>Methods in Enzymology</u>, Vol. 194, "Guide to Yeast Genetics and Molecular Biology", eds. C. Guthrie and G. Fink, Academic Press, Inc., Chap. 19, pp. 281-301 (1991).

Schedl et al., "A yeast artificial chromosome covering the tyrosinase gene confers copy number-dependent expression in transgenic mice", <u>Nature</u>, Vol. 362, pp. 258-261 (1993).

Strauss et al., "Germ line transmission of a yeast artificial chromosome spanning the murine a_1 (I) collagen locus", <u>Science</u>, Vol. 259, pp. 1904-1907 (1993).

Gilboa, E, Eglitis, MA, Kantoff, PW, Anderson, WF: Transfer and expression of cloned genes using retroviral vectors. BioTechniques 4(6):504-512, 1986.

Cregg JM, Vedvick TS, Raschke WC: Recent Advances in the Expression of Foreign Genes in *Pichia pastoris*, Bio/Technology 11:905-910, 1993

Culver, 1998. Site-Directed recombination for repair of mutations in the human ADA gene. (Abstract) Antisense DNA & RNA based therapeutics, February, 1998, Coronado, CA.

Huston et al, 1991 "Protein engineering of single-chain Fv analogs and fusion proteins" in Methods in Enzymology (JJ Langone, ed.; Academic Press, New York, NY) 203:46-88.

Johnson and Bird, 1991 "Construction of single-chain Fvb derivatives of monoclonal antibodies and their production in *Escherichia coli* in Methods in Enzymology (JJ Langone, ed.; Academic Press, New York, NY) 203:88-99.

Mernaugh and Mernaugh, 1995 "An overview of phage-displayed recombinant antibodies" in Molecular Methods In Plant Pathology (RP Singh and US Singh, eds.; CRC Press Inc., Boca Raton, FL) pp. 359-365.